Genome-wide scanning versus candidate gene approach in the genetic architecture of common diseases

In the Nature 7 June issue, researchers from the Welcome Trust Case Control Consortium (WTCCC) reported the findings of a large genome-wide association (GWA) study of 14,000 cases of common diseases and 3000 controls studies revealing the role of several loci in the genetic risk of common diseases [1].

Interestingly, almost simultaneously, researchers from the Diabetes Genetics Initiative (DGI) [2], the Finland–United States Investigation of NIDDM Genetics (FUSION) [3] and the Welcome Trust Case Control Consortium (WTCCC) [4] showed the results of other 3 large GWA studies in type 2 diabetes (T2D).

We wish to make a reflection about what we have learned from these new discoveries that could open powerful hints in the understanding of the pathophysiology of the diseases, particularly T2D, as pooling all the studies, there were 16,586 patients and 20,968 healthy controls, including the replication sets.

First, by pooling the results of all studies, the authors either confirmed or identified and confirmed the association of SNPs at three previously unknown loci such as *CDKN2B*, *CDKAL1* of *IGF2BP2* with the risk of T2D. Besides, the 3 consortia also made the effort of evaluating the contribution of variants in loci previously published as associated with the disease. For instance, the SNP rs13266634, a nonsynonymous variant in zinc transporter *SLC30A8* and the rs1801282, a nonsynonymous variant in the *PPARG* gene, were significantly associated when pooling all the data in spite of the fact that, in single genome scan (DGI), neither gene showed a positive signal for the aforementioned variants (p value: 0.92 and 0.83 respectively). Interestingly, the studies prove the worth of candidate gene studies as, for instance, the association results for the common *PPARG* Pro12Ala polymorphism were expressly investigated in the light of the preceding reports showing the association between the variant and enhanced insulin sensitivity and protection against T2D [5]. Thus, we wonder whether this SNP would have been carefully looked at even had no prior reports existed, as the data presentation reinforcing the previous published loci seems to be not unbiased regarding both the presumed information on protein function of the causal variant and the biological driven assumption. At that point, the whole efficiency of the genome-scans seems to be strongly shared between the exploratory- approach and the hypothesis-driven research.

Finally, we wish to emphasize that by making public the access to databases, consortia allow other investigators to search for association between candidate genes and T2D-related phenotypes. As an example we found, in a pilot study enrolling 1100 individuals that SNPs in the *CLOCK* (i.e. rs1554483 and rs6843722) were associated with

Nature Precedings : doi:10.1038/npre.2007.1438.1 : Posted 18 Dec 2007

hypertension (p value < 0.01, and 0.01, respectively). Interestingly, the same association was found by the WTCCC study (rs1554483, rs4580704, rs6843722 and rs4864548 with p<0.039, 0.014, 0.049 and 0.019). However, these findings were not mentioned by the authors because of the stringent p value criteria logically used by them in the face of a WGA.

Then WGA may prove to be more cost effective if databases are used for seeds of hypothesis-driven candidate gene approaches with p value cut-off higher than the stringent values imposed by genome correction, even at the expense of finding false positive associations. We found difficult to match data from different consortia since the format of the datasets are not homogenous and only WTCCC gives the complete information for all the diseases

Silvia Sookoian, MD, PhD, Clinical and Molecular Hepatology.

ssookoian@lanari.fmed.uba.ar

Carlos J Pirola, PhD, FAHA. Molecular Genetics and Biology of the Metabolic

Syndrome. **pirola.carlos@lanari.fmed.uba.ar**

Medical Research Institute-UBA-CONICET

Combatientes de Malvinas 3150

1427-Ciudad Autónoma de Buenos Aires

Argentina

REFERENCES

1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).

2. Saxena,R., *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-1336 (2007).

3. Scott,L.J., *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-1345 (2007).

4. Zeggini,E., *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-1341 (2007).

5. Altshuler,D., *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76-80 (2000).